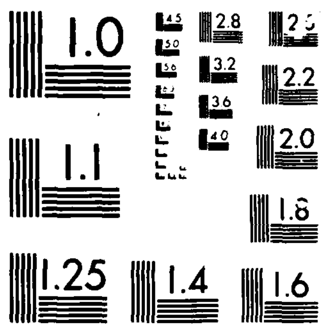
AD-A166 964 FILTERING AND CONTROL FOR WIDE BANDWIDTH NOISE DRIVEN SYSTEMS..(U) BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMICAL SYSTE.. H J KUSHNER ET AL. JAN 86 1/1

UNCLASSIFIED LCDS-86-8 AFOSR-TR-86-0258 DAAG29-84-K-0082 F/G 9/2 NL



MICROCOP CHART

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(2)

AD-A166 964

DTIC FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 86-0258	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Filtering and Control for Wide Bandwidth Noise Driven Systems		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) H. Kushner and W. Runggaldier		8. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0116
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lefschetz Center for Dynamical Systems, Division of Applied Mathematics, Brown University, Providence, RI 02912		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F, 2304, A1
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR/NM (Dr. Marc Jacobs) Bolling Air Force Base Washington, DC 20332		12. REPORT DATE January 1986
		13. NUMBER OF PAGES 53
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE APR 28 1986 S D B

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Much of modern stochastic control theory uses ideal white noise driven models (Itô equations). If the observed data is corrupted by noise, then the noise is usually assumed to be 'white Gaussian'. Typically, if the underlying models are linear, one uses a Kalman-Bucy filter to get an estimate of the

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

86 4 28 048

state, and then bases the control on this estimate. In practice, the noises are rarely 'white', and the reference signals and the systems are only approximations in some sense to a diffusion. Never-the-less, owing to lack of viable alternatives, one still uses the Kalman-Bucy filter, etc. Then the estimates are not optimal and, indeed, might be quite far from being optimal. Similarly for the corresponding control. (Examples are given to illustrate this.) The sense in which the estimates and/or control is useful need to be examined in order to justify the use of the commonly used procedure. The issue is much deeper than mere 'robustness' in the usual sense, since basic questions of interpretation of the results are involved. The paper deals with these questions. For the filtering problem where the signal is a 'near' Gauss-Markov process and the observation noise wide band, it is shown that the usual method is 'nearly optimal' with respect to a class of alternative data processors. This alternative class is rather natural and includes the data processors which one would normally want to use. It is unlikely that the class can be enlarged very much in general. The asymptotic (in time and bandwidth) problem is treated, as is the (much harder) conditional Gaussian case, and a case where the observations are non-linear. The basic techniques are those of weak convergence theory. Similar results are obtained for the combined filtering and control problem, where it is shown that good controls for the 'ideal' model are also good for the actual physical model, with respect to a natural class of alternative controls. The control problem over a finite interval as well as the average cost per unit time problem are considered.

(2)

AFOSR-TR- 86-0258

Lefschetz Center for Dynamical Systems

Division of Applied Mathematics

Brown University Providence RI 02912

Filtering and Control for Wide Bandwidth

Noise Driven Systems

by

H.J. Kushner and W. Runggaldier

January 1986 LCDS #86-8

Filtering and Control for Wide Bandwidth Noise Driven Systems

by

H.J. Kushner*
Lefschetz Center for Dynamical Systems
Division of Applied Mathematics
Brown University
Providence, Rhode Island 02912

and

W. Runggaldier**
Institute of Analysis
University of Padua
Padua, Italy
on leave at Brown University

January 1986

* This research supported in part by the Army Research Office under Grant #DAAG 29-84-K-0082; Office of Naval Research under Grant #N00014-83-K-0542; and the Air Force Office of Scientific Research under Grant #[redacted]F-AFOSR 81-0116-[redacted].

** This research supported in part by the Army Research Office under Grant #DAAG 29-84-K-0082 and the Office of Naval Research under Grant #N00014-83-K-0542.

Abstract

Much of modern stochastic control theory uses ideal white noise driven models (Itô equations). *If the observed data is corrupted by noise, then the* noise is usually assumed to be 'white Gaussian'. Typically, if the underlying models are linear, one uses a Kalman-Bucy filter to get an estimate of the state, and then bases the control on this estimate. In practice, the noises are rarely 'white', and the reference signals and the systems are only approximations in some sense to a diffusion. Never-the-less, owing to lack of viable alternatives, one still uses the Kalman-Bucy filter, etc. Then the estimates are not optimal and, indeed, might be quite far from being optimal. Similarly for the corresponding control. (Examples are given to illustrate this.) The sense in which the estimates and/or control is useful need to be examined in order to justify the use of the commonly used procedure. The issue is much deeper than mere 'robustness' in the usual sense, since basic questions of interpretation of the results are involved. The paper deals with these questions. For the filtering problem where the signal is a 'near' Gauss-Markov process and the observation noise wide band, it is shown that the usual method is 'nearly optimal' with respect to a class of alternative data processors. This alternative class is rather natural and includes the data processors which one would normally want to use. It is unlikely that the class can be enlarged very much in general. The asymptotic (in time and bandwidth) problem is treated, as is the (much harder) conditional Gaussian case, and a case where the observations are non-linear. The basic techniques are those of weak convergence theory. Similar results are obtained for the combined

filtering and control problem, where it is shown that good controls for the 'ideal' model are also good for the actual physical model, with respect to a natural class of alternative controls. The control problem over a finite interval as well as the average cost per unit time problem are considered.

Introduction

In much of modern control and filtering theory, one uses ideal white noise driven models of the following type, where $W_y(\cdot)$, $W_z(\cdot)$ and $W_x(\cdot)$ are standard Wiener processes, $u(\cdot)$ is a control, and b_z, σ_z, etc., are appropriate functions. We let $z(\cdot)$ denote a reference signal, $x(\cdot)$ the control system, $y(\cdot)$ the (integral of the) noise corrupted observation and $r_T(u)$ and $\gamma(u)$ the cost function.

(1.1) $$dz = b_z(z)dt + \sigma_z(z)dW_z$$

(1.2) $$dx = b_x(x,u)dt + \sigma_x(x)dW_x$$

(1.3) $$dy = h(x,z)dt + dW_y$$

(1.4) $$r_T(u) = \int_0^T E\ k(x(s), z(s), u(s))ds$$

(1.5) $$r(u) = \overline{\lim_{T\to\infty}}\ r_T(u)/T \quad .$$

Of course, the actual physical system, which we denote by $z^\epsilon(\cdot)$, $x^\epsilon(\cdot)$, $y^\epsilon(\cdot)$ (reference signal, control system, integral of the physical observation noise) is not of the form (1.1) - (1.3). The reference signal $z^\epsilon(\cdot)$ might be a 'near diffusion' - only approximately representable by (1.1), and the noise in the control and observation system would rarely be 'white'. But, typically, one somehow decides upon a suitable model (1.1) - (1.3), attempts to determine a good or 'nearly optimal' control for that model, and then applies this control to the actual physical system. In such a context, one must naturally question the value of the determined control for the 'physical' problem, as well as the value of the output of the filter (even for any 'nice' fixed control) for making estimates of functionals of the physical process $z^{\epsilon}(\cdot)$ which is approximated by $z(\cdot)$.

The filter output will rarely be even nearly optimal for use in making such estimates, and the control (based on the filter outputs) will rarely be 'nearly optimal'. Very little attention has been devoted to such problems - yet they are at the core of the problem of relevence of much theoretical work. An important theory of robustness has been developed [9], [10] - in which one tries to construct a filter in which the output is a continuous function of the input. The idea is that the model would be (1.1), (1.3), but with $W_y(\cdot)$ replaced by something else. Such robustness is very useful. But the very raising of such a robustness issue implies that the noises might *not* be white. If that is the case, what is the value of the filter (robust or not) - or of controls based on the filter output. Unless one is willing to assume more, there is no statistical interpertation of the output of such a filter. Furthermore, robustness must deal with the full control/filtering problem, correlation between the systems, the asymptotic (average cost per unit time problem), $z(\cdot)$, $x(\cdot)$ replaced by 'near diffusions', etc. We will deal with all these questions here, when the approximating system (1.1), (1.2) is linear - for which a fairly complete theory can be obtained.

Owing to the usual lack of 'near optimality' (for the physical system) of the filter and control which is obtained by using (1.1) - (1.3), one can only ask the question: with respect to which alternative filters ('data processors') or controls for the physical system are the chosen ones nearly optimal? It turns out, under quite broad condition, that this class of alternative filters and controls is quite large and very reasonable. Such results are essential, if the use of the ideal models (1.1) - (1.3) is to make sense in a large part of the applications.

The basic mathematical techniques used here are those of the theory of weak convergence of probability measures [1], [3], [4], a group of methods which are quite powerful for dealing with many difficult approximation problems in control and communication theory (and elsewhere) [1], [5] - [8], [14], and [15]. The basic questions of approximation here are closely related to those of the convergence of the sequence of physical processes $(z^\epsilon(\cdot), x^\epsilon(\cdot), y^\epsilon(\cdot))$ to the ideal model (1.1) - (1.3), as the 'bandwidth' of the driving noises (say, $1/\epsilon^2$) goes to infinity.

We begin with a discussion of the pure filtering problem. Here - for the case where the ideal model is linear - one would simply use the Kalman-Bucy filter for the ideal model - but whose input is the physical observation. Obviously, the filter does not usually yield the conditional distribution of the $z^\epsilon(t)$ given the data $y^\epsilon(s)$, $s \leq t$. In Section 2, we discuss some counter examples to illustrate the sort of difficulties which arise in such approximations, and in Section 3 the approximation theorem is given, together with the class of alternative data processors. Section 4 concerns the average filter error per unit time - or the errors for large time. We show that the filter output can be used to obtain estimates of a wide class of functionals of $z^\epsilon(\cdot)$, which are good with respect to a very broad and natural class of alternative estimators. The examples in Section 2 illustrate why they would not be 'nearly optimal' in general. In Sections 5 and 6, we treat the conditional Gaussian case, and a case where the observation is non-linear, and in Section 7, the non-linear observation case for large time. The power of the weak convergence approach should be amply evident in these sections. The conditional Gaussian case must be treated with some care, owing to the interaction between the wide bandwidth noise and the 'conditional Gaussian' coefficients. It is particularly important that any robustness or approximation theory be able to treat the large time - large bandwidth problem, and the conditional Guassian case and, at the moment, there seems to be no alternative to the weak convergence point of view for this.

The combined filtering and control problem is dealt with in Section 9. The optimal control for (1.1) - (1.3) will be nearly optimal for the physical system - in comparison with a large class of alternative controls. Appendix 1 contains some definitions concerning weak convergence. We will use the arrow $\Rightarrow$ to denote weak convergence. We have tried to formulate the models and results so that the paper is not burdened with a large amount of weak convergence theory or calculations - and so that available references can be used where possible. There are extensions in many directions: discrete parameter problems, impulsive control, etc., all treated very similarly to the treatment here.

2. 'Nearly' Optimal Linear Filtering: Formulation and Preliminaries

In the next few sections, we consider the following filtering problem: For each $\epsilon > 0$, $z^\epsilon(\cdot)$ is a *signal* process, $\xi_y^\epsilon(\cdot)$ is a 'wide-bandwidth' observation noise, the two are mutually independent and right continuous (with left hand limits) and the actual observation process is $\{\dot{y}^\epsilon(t), t \geq 0\}$:

$$\dot{y}^\epsilon(t) = H_z z^\epsilon(t) + \xi_y^\epsilon(t), \; y^\epsilon(0) = 0 \,. \tag{2.1}$$

The 'dependent' case can readily be handled. It's omitted in order to simplify the notation. Define $y^\epsilon(t) = \int_0^t \dot{y}^\epsilon(s)ds$ and $W_y^\epsilon(t) = \int_0^t \xi_y^\epsilon(s)ds$. Let $z(\cdot)$ be a Gauss- Markov process satisfying

$$dz = A_z z dt + B_z dW_z, \tag{2.2}$$

where $W_z(\cdot)$ is a standard Wiener process. The A_z, B_z, H_z are constant matrices, although they could be time-dependent in all parts, except those where $t \to \infty$.

We are concerned with the case where $\xi_y^\epsilon(\cdot)$ is 'nearly' white noise, and $z^\epsilon(\cdot)$ is 'nearly' a Gauss-Markov diffusion, and hence suppose that

$$(z^\epsilon(\cdot), W_y^\epsilon(\cdot)) \Rightarrow (z(\cdot), W_y(\cdot)) \quad \text{as} \quad \epsilon \to 0 \,, \tag{2.3}$$

where $W_y(\cdot)$ is a non-degenerate Wiener process. By the weak convergence and independence of $z^\epsilon(\cdot)$ and $\xi_y^\epsilon(\cdot)$, $W_z(\cdot)$ is independent of $W_y(\cdot)$. The weak limit of $\{y^\epsilon(\cdot)\}$ is $y(\cdot)$:

(2.4)
$$dy = H_z z dt + dW_y \ , \ y(0) = 0 \ .$$

Let $y \in R^k$, Euclidean k-space, and $z \in R^m$.

The actual physical system is 'fixed' and correspond to some small $\epsilon > 0$. The use of weak convergence here is just a way of embedding the *actual data* in a sequence - so that an approximation method can be used. We work with the 'near diffusion' $z^\epsilon(\cdot)$ and 'wide bandwidth' noise $\xi_y^\epsilon(\cdot)$. But to evaluate the filter that we design by using the ideal model (2.2), (2.4), *but with actual input* $\dot{y}^\epsilon(\cdot)$, the weak convergence method is very useful. W.p.1 convergence ideas are inappropriate in our context and would (in any case) restrict our flexibility. The 'distributional' information contained in the weak convergence is all that is needed, since the filters are evaluated by computing expectations of prediction errors. Similarly, the value of a control is evaluated via an expectation of a cost function - so only distributional information is needed.

We are interested in approximating the value of expectations of functions of $z^\epsilon(\cdot)$, conditioned on the data $y^\epsilon(\cdot)$. This is not easy. Except (and even then, rarely) for the special stationary and Gaussian cases of the classical Wiener theory, it is nearly impossible. Furthermore, if robustness is the issue, then we cannot restrict ourselves to Gaussian noise - since it itself is only an approximation to the physical processes.

For (2.2), (2.4), the classical Kalman-Bucy filter equations are

$$d\hat{z} = A_z\hat{z}dt + Q(t)\,[dy - H_z\hat{z}dt] \tag{2.5}$$
$$Q(t) = \Sigma(t)H_z'\,R_0^{-1}$$

$$\dot{\Sigma} = A_z\Sigma + \Sigma A_z' + B_zB_z' - \Sigma H'R_0^{-1}H\Sigma\,, \tag{2.6}$$

where R_0 = covariance matrix of observation 'noise' $W_y(1)$, which (w.l.o.g., by a simple rescaling) we set to I, unless mentioned otherwise. In practice, with physical wideband observation noise and the signal only a 'near' Gauss-Markov process, one normally uses (2.6) and the 'natural' adjustment of (2.5), namely

$$\dot{\hat{z}}^\epsilon = A_z\,\hat{z}^\epsilon + Q(t)\,[\dot{y}^\epsilon - H_z\hat{z}^\epsilon]\,. \tag{2.5$_{WB}$}$$

We want to know in what way the pair (2.5_{WB}), (2.6) makes sense. Typically, it is not an optimal - or even nearly optimal - filter, for the physical observation. But, as will be seen, it makes a great deal of sense and is quite appropriate in a specific but important way. One cannot ask whether it is 'nearly optimal' - but, rather, with respect to what class of alternative estimators is it 'nearly optimal' when estimating specific functionals of $z^\epsilon(\cdot)$. Weak convergence theory provides a natural tool for answering this question. Some of our results are related to these in [2], which concerns a non-linear filtering problem. But, for our specific case, it is possible to go further and get much more information fairly readily, and to treat the asymptotic (in time as well as in bandwidth) problem, various non-linear observation functions, the conditional Gaussian case, and the combined filtering and control problem; hence the overlap with [2] is very small.

Before proceeding, it is useful to consider several simple examples which illustrate the problems that we must contend with, particularly concerning the difference in the 'information' contained in the (integral of the) physical observation process $y^{\epsilon}(\cdot)$ and in the ideal (limit) $y(\cdot)$, and the *possible lack of continuity in the optimal estimators* as the noise bandwidth goes to ∞. Let (X_n,Y_n) be bounded real-valued random variables which converge in distribution (or even w.p.1) to (X,Y). Generally $E(X_n|Y_n) \not\to E(X|Y)$. X_n might be a physical signal and Y_n the physical observation, with the pair (X_n,Y_n) close in distribution to a much simpler pair (X,Y).
Consider the example where the lack of convergence is particularly evident:

Example 1. $X_n = X$, $Y_n = X/n$.

Example 2 illustrates a related pathology. If $Z_n = Z_n(Y)$, where Y is a random varible and $(Z_n,Y) \Rightarrow (Z,Y)$ (or even converges w.p.1). Then Z is *not* generally a function of Y.

Example 2. Let Y be uniformly distributed on $[0,1]$. Define $Z_n = nY$ for $0 \leq Y < 1/n$ and, in general, define $Z_n = (nY - k)$ on $k/n \leq Y < (k+1)/n$, $k = 0,1,\ldots,n-1$. Z_n is a 'sawtooth' function of Y. Also $(Z_n,Y) \Rightarrow (Z,Y)$ where Z is independent of Y, and both Z and Y are uniformly distributed on $[0,1]$. Clearly $E(Z_n|Y) \not\to E(Z|Y)$ in any sense.

Example 2 arises when we have a sequence of estimators, say $Z_{\epsilon}(y^{\epsilon}(\cdot))$, using the data $y^{\epsilon}(\cdot)$. Even if the pair $(Z_{\epsilon}(y^{\epsilon}(\cdot)), \ y^{\epsilon}(\cdot))$

converges to a pair $(Z, y(\cdot))$, the limit Z might *not* be a function of $y(\cdot)$. Here the limit is, in fact, independent of the data $y(\cdot)$. Similarly, for a control problem using noise corrupted data, say $y^\epsilon(\cdot)$. The limit control might be independent of the limit data!

Even though $W_y^\epsilon(\cdot) \Rightarrow W_y(\cdot)$, a non-degenerate Wiener process, $y^\epsilon(\cdot)$ might contain a *great deal* more information about $z^\epsilon(\cdot)$ than $y(\cdot)$ does about $z(\cdot)$. For an extreme case, consider

Example 3. Let t_i^ϵ, $i > 0$, be a strictly increasing sequence of real numbers for each ϵ, such that $t_i^\epsilon \xrightarrow{i} \infty$ and $\sup_i |t_{i+1}^\epsilon - t_i^\epsilon| \xrightarrow{\epsilon} 0$. Define $\Delta_i^\epsilon = t_{2i+1}^\epsilon - t_{2i}^\epsilon$, and for any $t > 0$, let $\sum_{t_i^\epsilon \leq t} \Delta_i^\epsilon \xrightarrow{\epsilon} 0$. Define a 'new' observation noise $\xi_y^\epsilon(\cdot)$ by *resetting* $\xi_y^\epsilon(t) = 0$ for $t \in [t_{2i}^\epsilon, t_{2i+1}^\epsilon)$, all i. The integral of the new $\xi_y^\epsilon(\cdot)$ still converges weakly to the Wiener process $W_y(\cdot)$. But $H_z z^\epsilon(\cdot)$ is known nearly exactly for small ϵ. There are even forms of this example for which the new $\xi^\epsilon(\cdot)$ is stationary.

These examples are admittedly pathological. But we are working with vague concepts such as 'wide bandwidth' observation noise, 'near' Gauss-Markov processes, and with integrals of the observation (as one always does in modern filtering theory). The examples do caution us to take considerable care. The examples showed that we might lose information in going to the limit. The following lemma (whose truth was first told to the author by T. Kurtz) shows the sense in which we never gain information on going to the limit - (i.e. noise bandwidth $\to \infty$).

Definition. For a set G, ∂G = (closure of G) minus (interior of G) = boundary of G. For a random variable Y, let $B(Y)$ be the minimal σ-algebra measuring Y, and let $I_G(Y)$ denote the indicator function of the set $\{\omega: Y \in G\}$.

Lemma 2.1. *Let* $(X_n,Y_n) \Rightarrow (X,Y)$ (X_n*-real valued*, Y_n *with values in* R^g). *Then*

$$\overline{\lim_n}\, E[X_n - E(X_n|Y_n)]^2 \leq E[X - E(X|Y)]^2 . \tag{2.7}$$

Remark. In Examples 1 to 3, the inequality is strict.

The proof is in Appendix 2. There is a similar result when Y_n is replaced by a (cadlag: right continuous with left hand limits) random process.

3. A Class of Data Processors (*Estimators*)

For the ideal filtering problem with data (2.2), (2.4), the optimal decision functions are functions of the estimates $\hat{z}(\cdot)$, $\Sigma(\cdot)$ since these completely determine the conditional distribution. There are no alternative (admissible) functions of the data $y(\cdot)$ which are better. This is *not so* with estimates based on $\Sigma(\cdot)$, $\hat{z}^\epsilon(\cdot)$ for the system $z^\epsilon(\cdot)$, $y^\epsilon(\cdot)$. We now define a large class of functions of the observed data $y^\epsilon(\cdot)$ with respect to which functions of $\hat{z}^\epsilon(\cdot)$, $\Sigma(\cdot)$ *are* 'nearly optimal' for small $\epsilon > 0$, and a large class of risk or cost functions. In order to know how good estimates based on $\hat{z}^\epsilon(\cdot)$, $\Sigma(\cdot)$ are for getting information on $z^\epsilon(\cdot)$, we need to specify both a class of (observation data dependent) alternative estimators - as well as a criterion of comparison; i.e., a cost function. We work with only one particular cost function - but the general idea and the natural extensions should be clear, and the method works with 'typical' cost functions.

Let $\mathcal{D}$ denote the class of measurable functions on $C[0,\infty]$, the space of real valued continuous functions on $[0,\infty)$ (with the topology of uniform convergence on bounded intervals), which are continuous w.p.1 relative to Wiener measure (hence, with repect to the measure of $y(\cdot)$). Let $\mathcal{D}_t$ denote the subclass which depends only on the function values up to time t. For arbitrary $F(\cdot) \in \mathcal{D}$ or in $\mathcal{D}_t$, we will use $F(y^\epsilon(\cdot))$ as an *alternative estimator* of a functional of $z^\epsilon(\cdot)$. The class is quite large, as will now be seen.

First, note that $\mathcal{D}$ contains all continuous functions and that the $\hat{z}(\cdot)$ of (2.5) can be written as a continuous function of the integral of the driving force $y(\cdot)$. [To see the latter point, solve (2.5) - in the form of a Wiener integral and do an integration by parts.] Thus, continuous functions of $\hat{z}^\epsilon(\cdot)$ are admissible estimators. Many important functionals are only continuous w.p.1 (relative to Wiener or $y(\cdot)$ measure). For some integer n, let A be a Borel set in R^{nk} with ∂A having zero Lebesque measure). Then [3], the function $I_A(x(t_1),...,x(t_n))$ is in $\mathcal{D}_t$ for any $t_1,...,t_n \leq t$, where $x(\cdot)$ denotes the canonical function in $C[0,\infty)$. Let $\tau(x(\cdot))$ denote the first time that a closed set A with a piecewise differential boundary is reached by $x(\cdot)$. Then the function with values $T \cap \tau(x(\cdot))$ is in $\mathcal{D}_T$ for any $T < \infty$. Thus, our alternative estimators can involve stopping times. This is essential in sequential decision problems or whenever the cost or risk function involves first entrance times of a function of $y(\cdot)$ into a decision set.

$\mathcal{D}$ and $\mathcal{D}_t$ do not contain 'wild' functions such as those involving differentiation. We consider $\mathcal{D}$ and $\mathcal{D}_t$ as a class of *data processors*. It seems to contain a large enough class for practical applications when the corrupting noise is 'white'. For the 'limit' (white observation noise, system $z(\cdot)$) problem, one would usually want processors that are continuous functions (w.p.1) of the data $y(\cdot)$. See the comments following the theorem statement below.

The following is one the main 'robustness' or 'approximation' results. For a function $q(z)$, we write (P_t^ϵ,q) for the integral of $q(z)$ with respect to the *Gaussian distribution* with mean $\hat{z}^\epsilon(t)$ and covariance $\Sigma(t)$ - the *ersatz conditional measure* of $z^\epsilon(\cdot)$. We let the $q(\cdot)$ and $F(\cdot)$ below be bounded, but the theorem holds if $\{(P_t^\epsilon,q)^2,\ q^2(z^\epsilon(t)),\ F^2(y^\epsilon(\cdot)\}$ is uniformly integrable.

Theorem 3.1. *Assume the conditions on* $z^\epsilon(\cdot)$, $W_y^\epsilon(\cdot)$ *of Section* 2. *Then* $(\hat{z}^\epsilon(\cdot), z^\epsilon(\cdot), W_y^\epsilon(\cdot)) \Rightarrow (\hat{z}(\cdot), z(\cdot), W_y(\cdot))$. *Let* $F(\cdot) \in \mathcal{D}_t$ *be bounded. and* $q(\cdot)$ *bounded continuous and real valued. Then (the limits all exist)*

$$\text{(3.1)} \qquad \lim_{\epsilon} E[q(z^\epsilon(t)) - F(y^\epsilon(\cdot))]^2$$

$$\geq \lim_{\epsilon} E[q(z^\epsilon(t)) - (P_t^\epsilon, q)]^2.$$

Remark. The theorem states that (for a small ϵ) the ersatz conditional distribution is 'nearly optimal' with respect to a specific (but broad) class of alternative estimators. The alternative class includes those that make sense to use when the corrupting noise is white. If the noise is wide band, then it might not make sense to exploit its detailed structure and use *other* 'better' estimators. Doing so might, in practical cases, cause processing errors and other (unmodelled) noise effects. We chose the estimate of the conditional mean at t in (3.1) for illustrative purposes. Many other cost or risk functionals could be considered; e.g., integrals of estimation errors - or the use of the estimates for control purposes (see below). The comment on stopping times in the paragraph above the theorem is useful for sequential estimation - where one stops when some function of the data first hits a decision set.

The assertion concerning the weak convergence is obvious, but necessary, since we need to know that the limit of the cited ϵ-triple represents a true filtering problem - with all three components, the system $z(\cdot)$, the filter $\hat{z}(\cdot)$ and the observation noise (integral) $W_y(\cdot)$. The result would not make sense if only 2 out of the 3 components converged.

Proof. By the weak convergence of the hypotheses and the w.p.1 continuity of $F(\cdot)$ we also have the weak convergence

$$\left[q(z^{\epsilon}(t)),\ F(y^{\epsilon}(\cdot)),\ (P_t^{\epsilon},q)\right] \Rightarrow$$

$$\left[q(z(t)),\ F(y(\cdot)),\ (P_t,q)\right],$$

where $(P_t,q) = \int q(z)dN(\hat{z}(t),\ \Sigma(t);dz)$, and $N(\hat{z},\Sigma;\cdot)$ is the normal distribution with mean $\hat{z}$ and covariance Σ. Thus, the left and right sides of (3.1) converge to, respectively,

$$E[q(z(t)) - F(y(\cdot))]^2,$$

$$E\left[q(z(t)) - E[q(z(t))\,|\,y(s),\ s \le t]\right]^2 . \tag{3.2}$$

Now, the proof follows from the fact that the second expression is no greater than the first, since the conditional expectation is the optimal estimator.

Q.E.D.

4. The Large Time Problem (*Large* t, *small* ϵ)

The filtering system often operates over a very long time interval. For the model (2.2), (2.4), one would then use the stationary filter and any acceptable method of analysis should be able to handle this 'large time' problem. But with the system $y^{\epsilon}(\cdot)$, $z^{\epsilon}(\cdot)$, *two limits are involved* since both $t \to \infty$ and $\epsilon \to 0$, and it is important that the results *not depend on how* $t \to \infty$ and $\epsilon \to 0$, and that the use of the stationary limit filter is justified. The weak convergene method is well set up to handle this problem.. For convenience we make some additional assumptions.

C4.1. A_z *is stable.* (A_z,H_z) *is observable and* (A_z,B_z) *controllable.*

By (C4.1), (2.6), has a unique, positive definite stable limit $\bar{\Sigma}$. The second part of (C4.2) is unrestrictive. It says simply that *increments* in $W^{\epsilon}(\cdot)$ behave 'close' to a Wiener process for small ϵ - no matter what t is.

C4.2. $\xi_y^{\epsilon}(t)$ *takes the form* $\xi_y^{\epsilon}(t) = \xi_y(t/\epsilon^2)/\epsilon$, *where* $\xi_y(\cdot)$ *is a right continuous second order stationary process with integrable covariance function* $R(\cdot)$. Also, if $t_\epsilon \to \infty$ as $\epsilon \to 0$, then $W_y^{\epsilon}(t_\epsilon+\cdot) - W_y^{\epsilon}(t_\epsilon) \Rightarrow W_y(\cdot)$.

Remark. The model (C4.2) is a common way of modelling wide bandwidth noise, and is used to simplify a calculation below, and to avoid the details involved with other models. Note in particular that if $S_y(w)$ is the spectral density of $\xi_y(\cdot)$, then $S_y(\epsilon^2\omega)$ is the spectral density of $\xi_y^{\epsilon}(\cdot)$. The

$S_y(\epsilon^2 w)$ converges to the spectral density of white noise as $\epsilon \to 0$, if $S_y(\cdot)$ is continuous at $w = 0$. It will be clear from the proof below that the condition can be considerably weakened. We also make the rather unrestrictive assumption that the initial time is not important and that the $z^\epsilon(\cdot)$ processes do not explode:

C4.3. *If* $\{z^\epsilon(t_\epsilon)\}$ *converges weakly to a random variable* $z(0)$ *as* $\epsilon \to 0$, *then* $z^\epsilon(t_\epsilon + \cdot) \Rightarrow z(\cdot)$ *with initial condition* $z(0)$. *Also*

$$\sup_{\epsilon,t} E|z^\epsilon(t)|^2 < \infty.$$

Consistency. In order that $\hat{z}(\cdot)$, $\Sigma(\cdot)$, be a filter for $z(\cdot)$, $y(\cdot)$, it is necessary that the *initial conditions* be *consistent*. Let $N(\hat{z},\Sigma;A)$ denote the probability that the normal random variable (with mean $\hat{z}$, and covariance Σ) takes values in the set A. By *consistency*, we mean that $P\{z(0) \in A | \hat{z}(0), \Sigma(0)\} = N(\hat{z}(0), \Sigma(0);A)$. One cannot choose the initial (random) conditions arbitrarily. It should be obvious that if $\Sigma(0) = \bar{\Sigma}$ and $(z(0), \hat{z}(0))$ are the *stationary* random variables for (stable) (2.2) and (2.5), then the initial conditions are consistent.

The question of consistency arises in our work since as $\epsilon \to 0$ and $t \to \infty$, we do not know a-priori what the *limits* of $(\hat{z}^\epsilon(t), z^\epsilon(t))$ are. When we study the asymptotics as $t \to \infty$ and $\epsilon \to 0$, we will start the filter at some large t_ϵ and the initial condition of the limit equations must be consistent for the problem to make sense. Fortunately, they will be consistent - so we will have a proper filter. This problem is considerably more difficult in the non-linear case. Theorem 4.1 is the 'large time' extension of Theorem 3.1. The question of consistency is either ignored in filtering - or else implicitly assumed; e.g., one cannot allow both $z(0) = 1$ w.p.1 and that $\hat{z}(0)$ has a normal distribution with a nonzero covariance.

Theorem 4.1. *Assume the conditions of Section* 2 *and* (C4.1) - (C4.3). *Let* $q(\cdot)$ *be bounded and continuous and let* $F(\cdot) \in \mathcal{D}_t$. *Define* $y^\epsilon(s) = 0$, *for* $s \leq 0$ *and define* $y^\epsilon(-\infty,t,\cdot)$ *to be the 'reversed' function - with values* $(0 \leq \tau < \infty)$ $y^\epsilon(-\infty,t;\tau) = y^\epsilon(t-\tau)$. *Then, if* $t_\epsilon \to \infty$ *as* $\epsilon \to 0$,

$$\{z^\epsilon(t_\epsilon + \cdot), \hat{z}^\epsilon(t_\epsilon + \cdot), W_y^\epsilon(t_\epsilon + \cdot) - W_y^\epsilon(t_\epsilon)\} \Rightarrow (z(\cdot), \hat{z}(\cdot), W_y(\cdot)) \tag{4.1}$$

satisfying (2.3), (2.5), *and* $z(\cdot)$, $\hat{z}(\cdot)$ *are stationary. Also* (3.1) *holds in the form*

$$\lim_{\epsilon,t} E\,[q(z^\epsilon(t)) - F(y^\epsilon(-\infty,t;\cdot))]^2 \geq \lim_{\epsilon,t} E[q(z^\epsilon(t)) - (P_t^\epsilon,q)]^2 . \tag{4.2}$$

The limit of (P_t^ϵ,q) *is the expectation with respect to the stationary* $(\hat{z}(\cdot), \bar{\Sigma})$ *system.*

Proof. Suppose that $\{\hat{z}^\epsilon(t), \epsilon > 0, t < \infty\}$ is tight [i.e., $\sup_{\epsilon,t} P\{|\hat{z}^\epsilon(t)| \geq N\} \xrightarrow{N} 0$]. Then, by the hypothesis, $\{\hat{z}^\epsilon(t), z^\epsilon(t), \epsilon > 0, t < \infty\}$ is tight and each subsequence of

$\{z^\epsilon(t_\epsilon+\cdot),\ \hat{z}^\epsilon(t_\epsilon+\cdot),\ W_y^\epsilon(t_\epsilon+\cdot) - W_y^\epsilon(t_\epsilon),\ t_\epsilon < \infty,\ \epsilon > 0\}$ has a weakly convergent subsequence with limit satisfying (2.2), (2.5). Choose a weakly convergent subsequence (with $t_\epsilon \to \infty$) also indexed by ϵ and with limit denoted by $z(\cdot)$, $\hat{z}(\cdot)$, $W_y(\cdot)$. Suppose, for the moment, that $z(\cdot)$, $\hat{z}(\cdot)$ is stationary. (Clearly, $\Sigma(t) \to \bar{\Sigma}$ as $t \to \infty$.) If all limits are stationary, then the subsequence is irrelevant since the stationary solution is unique. Also, since the initial conditions of $\hat{z}(\cdot)$ and $z(\cdot)$ are consistent under stationarity, $(\hat{z}(\cdot), \bar{\Sigma})$ is the optimal filter for $y(\cdot)$, $z(\cdot)$. Inequality (4.2) is a consequence of this and the weak convergence (by the argument used in Theorem 3.1.).

We next prove tightness of $\{\hat{z}^\epsilon(t),\ \epsilon > 0,\ t < \infty\}$, and then the stationarity will be proved. We have

$$\dot{\hat{z}}^\epsilon = [A_z - Q(t)H_z]\hat{z}^\epsilon + Q(t)\,\xi(t/\epsilon^2)/\epsilon + Q(t)H_z\, z^\epsilon(t)\ . \tag{4.3}$$

Let $\Phi(t,\tau)$ denote the fundamental matrix for $[A_z - Q(t)H_z]$. There are $K < \infty$, $\lambda > 0$ such that $|\Phi(t,\tau)| \leq K \exp - \lambda(t-\tau)$. We have

$$\hat{z}^\epsilon(t) = \Phi(t,0)z^\epsilon(t) + \int_0^t \Phi(t,\tau)\ Q(\tau)\,\xi(\tau/_\epsilon 2)d\tau/\epsilon$$
$$+ \int_0^t \Phi(t,\tau)Q(\tau)Hz^\epsilon(\tau)d\tau\ .$$

A straightforward calculation using (C4.2 - C4.3) and the change of variable $\tau/\epsilon^2 \to \tau$ in the first integral yields

$$E\,|\hat{z}^{\epsilon}(t)|^2 \leq \text{constant}\,(1 + E|z^{\epsilon}(0)|^2)\,,$$

giving the desired tightness.

To prove the stationarity of the limit of any weakly convergent subsequence, we need only show stationarity of the limit values $(z(0), \hat{z}(0))$ of the initial conditions $(z^{\epsilon}(t_{\epsilon}), \hat{z}^{\epsilon}(t_{\epsilon}))$. For this, we use a 'shifting' argument. Fix $T > 0$ and take a weakly convergent subsequence of (indexed also by ϵ, and with $t_{\epsilon} \xrightarrow{\epsilon} \infty$)

$$\{z^{\epsilon}(t_{\epsilon}+\cdot),\ \ \hat{z}^{\epsilon}(t_{\epsilon}+\cdot),\ \ W_y^{\epsilon}(t_{\epsilon}+\cdot) - W_y^{\epsilon}(t_{\epsilon}),\ \ z^{\epsilon}(t_{\epsilon}-T+\cdot),\ z^{\epsilon}(t_{\epsilon}-T+\cdot),$$
$$W_y^{\epsilon}(t_{\epsilon}-T+\cdot) - W_y^{\epsilon}(t_{\epsilon}-T)\}$$

with limit $\{z(\cdot),\ \hat{z}(\cdot),\ W_y(\cdot),\ z_T(\cdot),\ \hat{z}_T(\cdot),\ W_{y,T}(\cdot)\}$. Clearly, $\hat{z}_T(T) = \hat{z}(0)$ and $z_T(T) = z(0)$. We do not know what $\hat{z}_T(0)$ or $z_T(0)$ are - but, *uniformly in* T, they belong to a tight set (bounded in probability): I.e., owing to the tightness of $\{\hat{z}^{\epsilon}(t), z(t), \epsilon > 0, t < \infty\}$, for each $p > 0$, there is an $N_p < \infty$ such that $P\{|\hat{z}_T(0)| + |z_T(0)| \geq N_p\} < p$ for *all* T and limits of convergent subsequences. Write (where $W_{z,T}(\cdot)$ 'drives' the equation for dz_T)

$$z(0) = z_T(T) = (\exp A_z T) z_T(0) + \int_0^T \exp A_z(T-\tau)\cdot B_z dW_{z,T}(\tau)$$

$$\hat{z}(0) = \hat{z}_T(T) = (\exp\,[A_z - Q(\infty)H_z]T)\hat{z}_T(0)$$
$$+ \int_0^T \exp\,[A_z - Q(\infty)H_z](T-\tau)\cdot(dW_{y,T}(\tau) + H_z z_T(\tau)d\tau)$$

Since T is arbitrary and the set of all possible $\{z_T(0)\}$ is tight, the stability of A_z and $(A_z - Q(\infty)H_z)$ implies that $z(0)$ is the stationary random variable, hence $z(\cdot)$ is stationary. Similarly, the pair $(z(\cdot), \hat{z}(\cdot))$ is stationary.

Q.E.D.

Remark. There is no analog of Theorem 4.1 if A_z is unstable (if $t_\epsilon \to \infty$ as $\epsilon \to 0$), since the limit of $z^\epsilon(t_\epsilon)$ then makes no sense. If $z^\epsilon(\cdot)$ satisfied $\dot{z}^\epsilon = A_z z^\epsilon + B\xi_z^\epsilon(\cdot)$ for appropriate B and $\xi_z^\epsilon(\cdot)$ (such that the limit of $z^\epsilon(t_\epsilon+\cdot)$ is $z(\cdot)$), then we can show that

$$\lim_{\epsilon,t} E[z^\epsilon(t) - \hat{z}^\epsilon(t)]\,[z^\epsilon(t) - \hat{z}^\epsilon(t)]' = \bar{\Sigma} .$$

5. The Conditional Gaussian Problem.

We now consider the 'wide bandwidth' observation noise analog of the conditional Gaussian problem [12]. Let $q_i(\cdot)$, $i=1,2$, be bounded and continuously differentiable matrix valued functions with $q_2(x)q_2(x)' \geq \alpha I$ for some $\alpha > 0$. The signal $z^\epsilon(\cdot)$ and noise $\xi_y^\epsilon(\cdot)$ satisfy the conditions of Section 2, but the observation is of the 'wide bandwidth and conditional Gaussian type', where the coefficients are data dependent:

$$(5.1) \qquad \dot{y}^\epsilon = q_1(\hat{z}^\epsilon)z^\epsilon + q_2(\hat{z}^\epsilon)\xi_y^\epsilon(t) \ ,$$

where $\xi_y^\epsilon(t) = \xi_y(t/\epsilon^2)/\epsilon$ also satisfies (C4.2) and (the rather unrestrictive) (C5.1) below.

C5.1 $$E\left|\int_s^t du\left[E\left[\xi_y(u)\ \xi_y(s)' \mid \xi_y(v),\ v \leq 0\right] - R(u-s)\right]\right| \to 0$$

as s,t, $\to \infty$.

Define $R_0 = \int_{-\infty}^{\infty} R(u)du$. Formerly we used $R_0 = I$.

In (5.1), the $q_i(\cdot)$ can depend on the covariance $\Sigma^\epsilon(\cdot)$ given by (5.4) with no change in the results. The $q_i(\cdot)$ can also be more general functions of $y^\epsilon(\cdot)$ - as will be clear from the development. For simplicity, we use (5.1). (The 'correction terms' are more complicated in the general case. See remarks below.) Such conditional Guassian systems arise, for example, when one uses the observed data to orient or focus the observing mechanism, and the signal and noise strength depend on the orientation. The results of the previous sections are no longer directly applicable, since there is a 'correction term' due to the 'non-independence' of $\xi_y^\epsilon(t)$ and its coefficent $q_2(\hat{z}^\epsilon(t))$ in (5.1) - and similarly for related terms in the filter (5.3).

To prepare ourselves for setting up the correct filter equations, it is useful to anticipate the 'correction terms' and center the ϵ-filter appropriately so that the limit equations are the desired ones. Define the vector

$$F(\hat{z}, \Sigma, \xi) = \begin{bmatrix} q_2(\hat{z}) \\ \Sigma\, q_1{}'(\hat{z})[q_2(\hat{z})R_0q_2{}'(\hat{z})]^{-1}\, q_2(\hat{z}) \end{bmatrix} \xi$$

and $G = (G_1,\ldots,G_{k+m})$ by (recall that $y(t) \in R^k$ and $z(t) \in R^m$ and let F_{i,z_j} denote the derivative of F_i with respect to z_j)

$$G_i(\hat{z},\Sigma) = \int_0^\infty E \sum_j F_{i,\hat{z}_j}(\hat{z}, \Sigma, \xi_y(t))F_j(\hat{z}, \Sigma, \xi_y(0))dt \ .$$

Let $G^y(\hat{z},\xi)$ (resp., $G^z(\hat{z},\xi)$) denote the first k, (resp., the last m) components of $G(\hat{z},\xi)$. By Appendix 3, $G(\cdot)$ is the proper correction term for the $(y^\epsilon,\hat{z}^\epsilon)$ system, if $\hat{z}^\epsilon$ were defined by the appropriate 'conditional Gaussian' form of (2.5_{WB}).

Define the *centered observation* and *filter*

(5.2) $$\tilde{y}^\epsilon = y^\epsilon - G^y(\hat{z}^\epsilon, \Sigma^\epsilon)$$

$$(5.3)\qquad \dot{\hat{z}}^{\epsilon} = A_z \hat{z}^{\epsilon} - G^z(\hat{z}^{\epsilon}) + \Sigma^{\epsilon} q_1{}'(\hat{z}^{\epsilon}) [q_2(\hat{z}^{\epsilon})R_0 q_2{}'(\hat{z}^{\epsilon})]^{-1} [\dot{y}^{\epsilon} - q_1(\hat{z}^{\epsilon})\hat{z}^{\epsilon}]$$

$$(5.4)\qquad \dot{\Sigma}^{\epsilon} = A_z \Sigma^{\epsilon} + \Sigma^{\epsilon} A_z' + B_z B_z' - \Sigma^{\epsilon} q_1'(\hat{z}^{\epsilon}) [q_2(\hat{z}^{\epsilon})R_0 q_2'(\hat{z}^{\epsilon})]^{-1} q_1(\hat{z}^{\epsilon}) \Sigma^{\epsilon} .$$

(5.3) and (5.4) will be the proper filter for $z^{\epsilon}(\cdot)$, $y^{\epsilon}(\cdot)$, in the sense that the limit is the usual 'conditional Gaussian' filter and an analog of Theorem 3.1 or 4.1 can be proved. Define the system

$$(5.5)\qquad dz = A_z z\, dt + B_z dW_z$$

$$(5.6)\qquad dy = q_1(\hat{z})z\, dt + q_2(\hat{z})dW_y$$

$$(5.7)\qquad d\hat{z} = A_z \hat{z}\, dt + \Sigma q_1'(\hat{z})R_0 [q_2(\hat{z})R_0 q_2'(\hat{z})]^{-1}(dy - q_1(\hat{z})\hat{z}\, dt)$$

$$(5.8)\qquad \dot{\Sigma} = A_z\Sigma + \Sigma A_z' + B_z B_z' - \Sigma q_1'(\hat{z}) [q_2(\hat{z})R_0 q_2'(\hat{z})]^{-1} q_1(\hat{z})\Sigma .$$

Note that (5.7, 5.8) is the *optimal filter* for (5.5, 5.6), where $\mathrm{cov}W_y(t) = tR_0$, and $W_z(\cdot)$ and $W_y(\cdot)$ are independent and $\mathrm{cov}W_z(t) = tI$.

Theorem 5.1 is the appropriate analog of Theorem 3.1.

Theorem 5.1. *Assume the conditions of Section* 2, (C4.2) *and* (C5.1). *Let the system* (5.5) - (5.8) *have a unique solution (in the sense of distributions) for each initial condition. Then*

$$\{z^\epsilon(\cdot),\ \hat{z}^\epsilon(\cdot),\ W_y^\epsilon(\cdot),\ \tilde{y}^{\,\epsilon}(\cdot),\ \Sigma^\epsilon(\cdot)\} \Rightarrow \qquad (z(\cdot),\ \hat{z}(\cdot),\ W_y(\cdot).\ y(\cdot),\ \Sigma(\cdot))\ , \tag{5.9}$$

where $z(\cdot)$ *and* $W_y(\cdot)$ *are mutually independent and cov* $W_y(t) = tR_0$. *Also* (3.1) *holds.*

Proof. (3.1) is a consequence of the weak convergence, and the weak convergence is a consequence of the results in Appendix 3.

Remark. The processes in (5.2) and (5.3) were centered so that the weak limits (5.7), (5.8) would be the correct filter for the limit system (5.5), (5.6). If we had not centered, then the limit of (the uncentered) $y^\epsilon(\cdot)$ equation would contain an additional drift term which would not be compensated for by the correction term in the limit of the uncentered $\hat{z}^\epsilon(\cdot)$ sequence; thus, the limit process $(\hat{z}(\cdot),\ \Sigma(\cdot))$ would not necessarily be a filter for the $(z(\cdot),\ y(\cdot))$ process.

Note that the correction (centering) terms involve first derivatives of $q_1(\cdot)$ and $q_2(\cdot)$, (although, via the centering, the limit does not involve the derivatives). This can lead to some unfortunate and generally ignored difficulties. Suppose that we can choose the $q_i(\cdot)$ and that we choose them to optimize some cost criterion. We can't do the optimization with the $(y^\epsilon,\hat{z}^\epsilon,\Sigma^\epsilon)$ system because that would be computationally impossible - but we can (in principle) with the limit system. But, unless the resulting 'control' $(q_i(\cdot),\ i=1,2)$ is continuously differentiable, it cannot be used, since the correction terms involve derivatives. In fact, it is not clear whether or not there is a weak convergence result for non-differentiable $q_i(\cdot)$. Similar problems arise wherever the coefficient of a 'wide bandwidth' noise process depends on a 'control'. If the $q_i(\cdot)$ depended on the $y^\epsilon(\cdot)$ or $\tilde{y}^\epsilon(\cdot)$ in a different (but 'smooth') way - other than via $(\hat{z}^\epsilon,\Sigma^\epsilon)$, there will usually be a (even more complicated) correction term. But its form can be worked out by the methods of weak convergence theory.

6. Nonlinear Observations

The ideas of the previous sections (and Section 8) are useful for problems which have a partly non-linear structure, but where the 'limit' system is linear. We now develop this for one special but important case. Many filtering or communcation systems use limiters on the input for purposes of increasing robustness or for 'linear dynamic range' reasons, when the power of the input can vary over a large range. The input is put through a 'hard' limiter - then followed by a linear filter, whose purpose is to reconstruct the input. Such systems have been of great interest in communication theory. See [13], one of the first attempts to systematically analyze such a system.We treat one case - where the observation is scalar valued and is

$$(6.1) \qquad \dot{y}^{\epsilon} = k(H_z z^{\epsilon} + \xi_y^{\epsilon}(t))/\epsilon, \quad k(x) = \mathrm{sign}(x), \quad y^{\epsilon}(0) = 0.$$

The $1/\epsilon$ is a normalizing term and can be put anywhere in the filter system - as long as the system is linear. The normalization might or might not be used in practice. The qualitative results will remain the same - but the average power in the *unnormalized* observation goes to zero as the bandwidth of $\xi_y^{\epsilon}(\cdot)$ goes to ∞. A similar development (with the same results) can be carried out with the use of a 'soft' limiter; i.e., $k(x) = \mathrm{sign}(x)$ for $|x| > c > 0$, $k(x) = x/c$ for $|x| < c$, and also if $k(\cdot)$ is vector valued.

We use

C6.1. $\xi_y^\epsilon(t) = \xi_y(t/\epsilon^2)/\epsilon$, *where* $\xi_y(\cdot)$ *is a component of a stationary Gauss-Markov process whose correlation function goes to zero as* $t \to \infty$ *(hence* $\to 0$ *exponentially).*

Write $E(\xi_y(t))^2 = \sigma_0^2$. Then the average of 6.1) over the noise ξ_y^ϵ is

(6.2) $$\left(\frac{2}{\pi\sigma_0^2}\right)^{\frac{1}{2}} H_z\, z^\epsilon(t) + \delta_\epsilon$$

where $\delta_\epsilon \to 0$ as $\epsilon \to 0$, uniformly for $z^\epsilon(t)$ in any bounded set. In preparation for the approximation result, define the systems

(6.3) $$dz = A_z z\, dt + B_z dW_z$$

(6.4) $$dy = \left(\frac{2}{\pi\sigma_0^2}\right)^{\frac{1}{2}} H_z z\, dt + 2J_0^{\frac{1}{2}}\, dW_y$$

(6.5) $$d\hat{z} = A_z \hat{z}\, dt + Q(t)[dy - \left(\frac{2}{\pi\sigma_0^2}\right)^{\frac{1}{2}} H_z \hat{z}\, dt]$$

(6.6) $$\dot{\Sigma} = A_z \Sigma + \Sigma A_z' + B_z B_z' - \Sigma H_z' H_z \Sigma \left(\frac{2}{\pi\sigma_0^2}\right)\left(\frac{1}{4J_0}\right)$$

(6.7) $$Q(t) = \Sigma(t) H_z' \left(\frac{2}{\pi\sigma_0^2}\right)^{\frac{1}{2}} \frac{1}{4J_0}$$

(6.8) $$J_0 = \frac{1}{\pi}\int_0^\infty \sin^{-1} \rho(t)\, dt ,$$

where $\rho(\cdot)$ is the correlation function of $\xi_y(\cdot)$. Define $\hat{z}^\epsilon(\cdot)$ by

(6.9) $$\dot{\hat{z}}^\epsilon = A_z \hat{z}^\epsilon + Q(t)[\dot{y}^\epsilon - \left(\frac{2}{\pi\sigma_0^2}\right)^{\frac{1}{2}} H_z \hat{z}^\epsilon] .$$

Equations (6.5) to (6.8) represent the Kalman-Bucy filter for the system (6.3), (6.4). Equations (6.6), (6.9) represent the filter which one would normally use for the system $(z^{\epsilon}(\cdot), y^{\epsilon}(\cdot))$, and whose use we must justify. The justification is by Theorem 6.1.

Theorem 6.1. *Assume the conditions of Section* 2 *and* (C6.1). *Then*

$$\{z^{\epsilon}(\cdot), \hat{z}^{\epsilon}(\cdot), y^{\epsilon}(\cdot)\} \Rightarrow (z(\cdot), \hat{z}(\cdot), y(\cdot)) . \tag{6.10}$$

and $W_y(\cdot)$ *is independent of* $z(\cdot)$. *Also,* (3.1) *holds.*

Remark. The power of the weak convergence methods is well illustrated by the relative ease of getting this result. The problem is very hard - due to the nautre of the nonlinearity, and alternative approaches to even a small part of the analysis (e.g., as in the classical work [13]) are very involved.

Proof. The proof of the weak convergence follows from that in [14], or [1], Chapter 9.3], and (3.1) follows from the weak convergence, exactly as in the proof of Theorem 3.1. Actually, the proofs in [14], [1] use a signal $s(\cdot)$ which does not depend on ϵ, but the proofs would be essentially unchanged if the actual ϵ-dependent signal $z^{\epsilon}(\cdot)$ were used instead.

7. The Large Time Problem: Nonlinear Observations

We now do the analog of the 'ergodic' case of Theorem 4.1 for the nonlinear observation problem of Section 6 for the case where the system of Section 6 is in operation for a long time.

Theorem 7.1. *Assume the conditions of Theorem* (6.1) *and* (C4.1) *to* (C4.3). *Then the conclusions of Theorem* 4.1 *hold, where* $z(\cdot)$, $\hat{z}(\cdot)$ *and* $\Sigma(\cdot)$ *are the stationary systems for* (6.3) to (6.7).

Remark on the proof. By the method of proof of Theorem 4.1, and the result of Theorem 6.1, in order to carry the proof of Theorem 4.1 to the present case, we need only show tightness of $\{\hat{z}^{\epsilon}(t), \epsilon > 0, t < \infty\}$. Due to the non-linearity of the observation, it is no longer possible to do it directly, as in Theorem 4.1, and a 'perturbed Liapunov function' method will be employed [1, Chapter 6]. Those methods are useful for getting stability-type results for 'wideband noise' driven or 'near' Markov processes, results that are generally hard to get. Such results are essential for the asymptotic (large t) analysis. The development will be essentially self contained, but the interested reader will find a fuller discussion and other applications in [1]. Related perturbed Liapunov function methods are used in [15].

Proof Part 1. We show only the above mentioned tightness. For use below, we first evaluate the expression

$$\text{(7.1)} \qquad K^{\epsilon}(t) = \frac{1}{\epsilon} \int_t^{\infty} \left[E_t^{\epsilon} k(H_z z^{\epsilon}(s) + \xi_y^{\epsilon}(s)) - \bar{E}_t^{\epsilon} k(H_z z^{\epsilon}(s) + \xi_y^{\epsilon}(s))\right] ds,$$

where E_t^{ϵ} denote the expectation conditioned on $\{\xi_y^{\epsilon}(u), z^{\epsilon}(u), u \leq t\}$, and $\bar{E}_t^{\epsilon}$ denotes the expectation conditioned on $\{z^{\epsilon}(u), u \leq t\}$ and under the assumption that $\xi_y^{\epsilon}(s)$ is the *stationary* random variable. Let $\xi_y(t) = H_0 \bar{\xi}_y(t)$ for some matrix H_0, where $\bar{\xi}_y(\cdot)$ is the Gauss-Markov process cited in (C6.1). Note that there are $\lambda > 0$ and $c_0 < \infty$ such that

$$\text{(7.2)} \qquad |\text{variance [stationary } \xi_y(t)] - \text{variance } [\xi_y(t) | \bar{\xi}_y(0) = 0]| \leq c_0 \exp - \lambda t ,$$

$$|E[\xi_y(t) | \bar{\xi}_y(0)]| \leq (c_0 \exp - \lambda t) |\bar{\xi}_y(0)| .$$

Changing scale $s/\epsilon^2 \to s$ in (7.1) and multiplying the arguments of $k(\cdot)$ by ϵ yields

$$\text{(7.3)} \qquad K^{\epsilon}(t) = \epsilon \int_{t/\epsilon^2}^{\infty} \left[E_t^{\epsilon} k(\epsilon H_z z^{\epsilon}(\epsilon^2 s) + \xi_y(s)) - \bar{E}_t^{\epsilon} k(\epsilon H_z z^{\epsilon}(\epsilon^2 s) + \xi_y(s))\right] ds.$$

For large initial conditions (at time t/ϵ^2) $\bar{\xi}_y(t/\epsilon^2)$, (7.1) might be large. For $|\bar{\xi}_y(t/\epsilon^2)| \geq 1$ and $s - t/\epsilon^2 \geq 0(\log |\bar{\xi}_y(t/\epsilon^2)|)$, the conditional mean of $\xi_y(s)$ (given $\bar{\xi}_y(t/\epsilon^2)$) will be $0(1)$. Thus, we can write

$$|K^{\epsilon}(t)| = 0(\epsilon)\left[|\bar{\xi}_y(t/\epsilon^2)| + 1\right] + |K^{\epsilon}(t)| I_{\{|\bar{\xi}_y(t/\epsilon^2)| \leq 1\}} .$$

We now deal with initial values $\bar{\xi}_y(t/\epsilon^2) = 0(1)$. Let $N(a,b)$ denote a normally distributed *random variable* with mean a and variance b. In evaluating the expression

$$E_t^\epsilon\, k(\epsilon H_z z^\epsilon(\epsilon^2 s) + \xi_y(s)) - \bar{E}_t^\epsilon\, k(\epsilon H_z z^\epsilon(\epsilon^2 s) + \xi_y(s)) \ , \epsilon^2 s \geq t,$$

we can replace the conditional expectations by expectations over $\xi_y(s)$ only, where the first $\xi_y(s)$ can be taken to be $N(\delta_1, \sigma_0^2\,(1-\delta_2))$ and the second can be taken to be $N(0,\sigma_0^2)$, where $\delta_i \to 0$ exponentially as $(s - t/\epsilon^2) \to \infty$. For notational simplicity, set $\sigma_0^2 = 1$.

For small $\delta_i > 0$ and $z > 0$ (with a similar development for $z < 0$),

$$|P\{|N(\delta_1, 1-\delta_2)| \leq \epsilon z\} - P\{|N(0,1)| \leq \epsilon z\}|$$

$$\leq |P\{|N(0,1-\delta_2)| \leq \epsilon z\} - P\{|N(0,1)| \leq \epsilon z\}| + 0(\delta_1)$$

$$\leq |P\{|N(0,1)| \leq \epsilon z(1+2\delta_2)\} - P\{|N(0,1)| \leq \epsilon z\}| + 0(\delta_1)$$

$$\leq 2P\{\epsilon z \leq N(0,1) \leq \epsilon z(1+2\delta_2)\} + 0(\delta_1) = 0(\epsilon z \delta_2) + 0(\delta_1) \ .$$

Putting these estimates into (7.3) and using the cited fact that $\delta_i \to 0$ exponentially, for some $\lambda > 0$ we have

$$(7.4) \qquad |K^\epsilon(t)| \leq 0(\epsilon)\left[|\bar{\xi}_y(t/\epsilon^2)| + 1\right]$$

$$+ 0(\epsilon^2) \int_{t/\epsilon^2}^\infty \bar{E}_t^\epsilon |z^\epsilon(\epsilon^2 s)| \exp - \lambda(s-t/\epsilon^2) \cdot ds$$

$$= K_1^\epsilon(t)$$

$$E|K_1^{\epsilon}(t)|^2 = 0(\epsilon^2) \tag{7.5}$$

Part 2. Write (6.9) as

$$\dot{\hat{z}}^{\epsilon} = \left[A_z - Q(t)H_z \left(\frac{2}{\pi\sigma_0^2}\right)^{1/2}\right]\hat{z}^{\epsilon} + Q(t)\, k^{\epsilon}(t)/\epsilon, \tag{7.6}$$

where we use

$$k^{\epsilon}(t) = k(H_z z^{\epsilon}(t) + \xi_y^{\epsilon}(t)).$$

For our stability argument, $Q(\infty)$ can be used in lieu of $Q(t)$ in (7.6) (justified by a 'perturbation argument, which we omit). Define

$$A_1 = [A_z - Q(\infty)H_z\left(\frac{2}{\pi\sigma_0^2}\right)^{1/2}],$$

and let $P > 0$ be such that $A_1 P + PA_1' = -C < 0$. We start with the Liapunov function $\hat{z}'P\hat{z} = V(\hat{z})$, and then 'perturb' it. See Appendix 4 for the definition of the operator $\hat{A}^{\epsilon}$ below. (It is essentially a 'differentiation' operator.) By Appendix 4, we have

$$\begin{aligned} \hat{A}^{\epsilon}\, V(\hat{z}^{\epsilon}(t)) &= \dot{V}(\hat{z}^{\epsilon}(t)) \\ &= -\hat{z}^{\epsilon\,\prime}(t)\; C\; \hat{z}^{\epsilon}(t) + 2\hat{z}^{\epsilon}(t)'PQ(\infty)k^{\epsilon}(t)/\epsilon. \end{aligned} \tag{7.7}$$

The second term on the right of (7.7) is not dominated by the first, and we 'perturb' the Liapunov function in order to 'control' the bad term. Define the perturbation

$$V_1^\epsilon(t) = \frac{2\hat{z}^{\epsilon\prime}(t)P}{\epsilon} \int_t^\infty Q(\infty)\left[E_t^\epsilon k^\epsilon(s) - \bar{E}_t^\epsilon k^\epsilon(s)\right]ds \ .$$

By Part 1, $|V_1^\epsilon(t)| = 0(1)|\hat{z}^\epsilon(t)|K_1^\epsilon(t)$ and by Schwarz's inequality and (7.4),

$$\text{(7.8)} \qquad E|V_1^\epsilon(t)| = 0(\epsilon)\ E^{1/2}|\hat{z}^\epsilon(t)|^2$$

$$= 0(\epsilon)\ \left[1 + E|\hat{z}^\epsilon(t)|^2\right] .$$

Also, we can readily show that

$$\text{(7.9)} \qquad \hat{A}^\epsilon V_1^\epsilon(t) = -\frac{2\hat{z}^\epsilon(t)}{\epsilon}\ PQ(\infty)\left[k^\epsilon(t) - \bar{E}_t^\epsilon k^\epsilon(t)\right]$$

$$+ \frac{2\dot{\hat{z}}^{\epsilon\prime}(t)P}{\epsilon} \int_t^\infty Q(\infty)\left[E_t^\epsilon k^\epsilon(s) - \bar{E}_t^\epsilon k^\epsilon(s)\right]ds \ .$$

Recall that the $\bar{E}_t^\epsilon k^\epsilon(s)$ is the expectation over the stationary $\xi_y(s)$ only: i.e., the conditioning data is just $z^\epsilon(\cdot)$. It can be shown that

$$E|\hat{z}^\epsilon(t)|\left|\frac{\bar{E}_t^\epsilon k^\epsilon(t)}{\epsilon}\right| = 0(1)E^{1/2}|\hat{z}^\epsilon(t)|^2 E^{1/2}|z^\epsilon(t)|^2 \ .$$

By substituting (6.9) for $\dot{\hat{z}}$ and using our bound on the integral, the last term on the right of (7.9) is bounded by

$$0(1)\ \left[|\hat{z}^\epsilon(t)| + 1\right]K_1^\epsilon(t) + 0(1)\ K_1^\epsilon(t)/\epsilon \ .$$

Define the perturbed Liapunov function $V^\epsilon(t) = V(\hat{z}^\epsilon(t)) + V_1^\epsilon(t)$. Putting the estimates together, evaluating $\hat{A}^\epsilon V^\epsilon(t)$ via (7.7) and (7.9) and cancelling the common $1/\epsilon$ terms (with opposite signs) in (7.7) and (7.9) and using (See (A4.1) in Appendix 4)

$$EV^{\epsilon}(t) = EV^{\epsilon}(0) + \int_0^t E\hat{A}^{\epsilon} V^{\epsilon}(s)ds$$

and the bound on $E|V_1^{\epsilon}(t)|$ yields

$$\begin{aligned} EV(\hat{z}^{\epsilon}(t)) &\leq \text{constant} + 0(\epsilon)\left[1 + E|\hat{z}^{\epsilon}(t)|^2\right] \\ &+ \int_0^t (\text{constant})\, ds - \int_0^t E\hat{z}^{\epsilon\prime}(s)C\hat{z}^{\epsilon}(s)ds \\ &+ (\text{constant}) \int_0^t E(|\hat{z}^{\epsilon}(s)| + 1)K_1^{\epsilon}(s)/\epsilon\, ds \\ &+ (\text{constant}) \int_0^t E^{\frac{1}{2}}|\hat{z}^{\epsilon}(s)|^2 E^{\frac{1}{2}}|z^{\epsilon}(s)|^2 ds\,. \end{aligned} \tag{7.10}$$

Using the inequality $|ab| \leq a^2/c + cb^2$ for any $c > 0$ and (7.5) in the last two integrals of (7.10) yields, for some constants $c_i > 0$ and $c > 0$,

$$\begin{aligned} E|\hat{z}^{\epsilon}(t)|^2 &\leq c_1(1+t) - c_0 \int_0^t E|\hat{z}^{\epsilon}(s)|^2 ds \\ &+ \frac{c_2}{c} \int_0^t E|\hat{z}^{\epsilon}(s)|^2 ds + c\, c_3\, t\,. \end{aligned} \tag{7.11}$$

By letting $c_2/c < c_0$, (7.11) implies that

$$\sup_{\epsilon,t} E|\hat{z}^{\epsilon}(t)|^2 < \infty\,. \tag{7.12}$$

Finally, (7.12) is equivalent to the required tightness of $\{\hat{z}^{\epsilon}(t),\ \epsilon > 0,\ t < \infty\}$.

Q.E.D.

8. The Filtering and Control Problem: Finite Time Case.

The ideas and results of the foregoing sections can be extended to the combined filtering and control problem. The issues remain essentially the same. As seen in the previous sections, the use of the (suitably adjusted when necessary) Kalman-Bucy filter for the wide bandwidth observation noise and 'near Gauss-Markov' signal might be very far from optimal, but it is 'nearly optimal' with respect to a large and reasonable class of alternative data processors. For the combined filtering and control case, the issue is more complicated. The control system will be driven by wide bandwidth noise as well, and neither the system nor the reference signals would be Markov. Suppose that (as is usually the case) one obtains a control (optimal or not) based on the usual ideal white noise driven limit model. This control will be a function of the outputs of the filters, and one must question the value of applying this to the actual wide bandwidth noise system.

Consider the following linear control and filtering problem: Let $z^\epsilon(\cdot)$ denote the physical reference signal and let $z^\epsilon(\cdot) \Rightarrow z(\cdot)$ as $\epsilon \to 0$ where, as in the previous sections, $z(\cdot)$ satisfies

$$dz = A_z z\, dt + B_z dW_z . \tag{8.1}$$

Let the control system be (for constant matrices A_x, D_x, B_x, H_x) defined by

$$\dot{x}^\epsilon = A_x x^\epsilon + D_x u + B_x \xi_x^\epsilon , \tag{8.2}$$

with observations $\dot{y}^\epsilon(\cdot)$, where

$$(8.3)\qquad \begin{bmatrix} \dot{y}_z^\epsilon \\ \dot{y}_x^\epsilon \end{bmatrix} \equiv \dot{y}^\epsilon = \begin{bmatrix} H_z z \\ H_x x \end{bmatrix} + \xi^\epsilon , \quad y^\epsilon \in R^g , \ y^\epsilon(0) = 0 ,$$

where the three processes $\int_0^t \xi^\epsilon(s)ds \equiv W^\epsilon(t)$, $\int_0^t \xi_x^\epsilon(s)ds \equiv W_x^\epsilon(t)$ and $z^\epsilon(\cdot)$ are mutually independent, and $W^\epsilon(\cdot) \Rightarrow W(\cdot)$, $W_x^\epsilon(\cdot) \Rightarrow W_x(\cdot)$, standard Wiener processes. Thus $\xi_x^\epsilon(\cdot)$ and $\xi^\epsilon(\cdot)$ are *wide bandwidth noise processes.* Correlations among these processes can also be handled, at the expense of a more complex notation.

Define the filters and limit system:

$$(8.4)\qquad \begin{bmatrix} \dot{\hat{x}}^\epsilon \\ \dot{\hat{z}}^\epsilon \end{bmatrix} = \begin{bmatrix} A_x \hat{x}^\epsilon \\ A_z \hat{z}^\epsilon \end{bmatrix} + \begin{bmatrix} D_x u \\ 0 \end{bmatrix} + Q(t) \left[\dot{y}^\epsilon - \begin{bmatrix} H_x \hat{x}^\epsilon \\ H_z \hat{z}^\epsilon \end{bmatrix} \right],$$

$$(8.5)\qquad dy = \begin{bmatrix} H_x x \\ H_z z \end{bmatrix} dt + dW = H \begin{bmatrix} x \\ z \end{bmatrix} + dw$$

$$(8.6)\qquad d\begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} A_x \hat{x} \\ A_z \hat{z} \end{bmatrix} dt + \begin{bmatrix} D_x u \\ 0 \end{bmatrix} Q(t) \left[dy - \begin{bmatrix} H_x \hat{x} \\ H_z \hat{z} \end{bmatrix} dt \right],$$

$$(8.7)\qquad dx = A_x x dt + D_x u dt + B_x dW_x ,$$

with the obvious associated Ricatti equation for the conditional covariance $\Sigma(\cdot)$ of $(x(\cdot), z(\cdot))$. Here $Q(t) = \Sigma(t)H'[\mathrm{cov}\ W(1)]^{-1}$. Equation (8.4) will be the filter for $(x^\epsilon(\cdot), z^\epsilon(\cdot))$ with data $y^\epsilon(\cdot)$, and (8.6) is the filter for (8.5), (8.7). The cost functions for the control problem are

(8.8) $$R^{\epsilon}(u) = \int_0^T E\ r(x^{\epsilon}(t),\ z^{\epsilon}(t),\ u(t))dt,$$

(8.9) $$R(u) = \int_0^T E\ r(x(t),\ z(t),\ u(t))dt,$$

for bounded and continuous $r(\cdot,\cdot,\cdot)$, and some $T < \infty$.

The controls take values in a compact set U, and we let (see related definition of $\mathcal{D}$ and $\mathcal{D}_t$ in Section 3) $\mathcal{H}$ denote the set of U-valued measurable (ω,t) functions on $C^{g}[0,\infty) \times [0,\infty)$ which are continuous w.p.1. relative to Wiener measure. Let $\mathcal{H}_t$ denote the subclass which depends only on the function values up to time t. We view functions in $\mathcal{H}$ as the data dependent controls with value $u(y(\cdot),t)$ at time t and data $y(\cdot)$. Let $\overline{\mathcal{H}}$ denote the subclass of functions $u(\cdot,\cdot) \in \mathcal{H}$ such that $u(\cdot,t) \in \mathcal{H}_t$ for all t and with the use of control $u(y^{\epsilon}(\cdot),\cdot)$ (resp., $u(y(\cdot),\cdot)$), (8.2) and (8.4) (resp., (8.6), (8.7)) has a unique solution *in the* sense of distibutions. These $u(y^{\epsilon}(\cdot),\cdot)$ and $u(y(\cdot),\cdot)$ are the admissible controls.

Commonly, one tries to use the model (8.5), to (8.7) to get a (nearly) optimal control for cost (8.9). This control would, in practice, actually be applied to the 'physical' system (8.2), (8.4), with actual cost function (8.9). Such controls would normally not be 'nearly' optimal in any strict sense for the physical systems and questions arise which are similar to those posed for the pure filtering problem: in particular, with respect to what class of *comparison controls* is such a control 'nearly optimal'. Again, weak convergence theory can provide some answers, although the problem is considerably more difficult, and the results less satisfactory.

Straightforward weak convergence arguments (using only the assumed weak convergence of the 'driving $W^\epsilon(\cdot)$, $W_x^\epsilon(\cdot)$ processes', and the uniqueness of the limit can be used to prove Theorem 8.1. Let $\mathcal{M}$ denote the class of U-valued continuous functions $u(\cdot,\cdot,\cdot)$ such that with use of control with value $u(\hat{x}(t), \hat{z}(t),t)$ at time t, (8.6), (8.7), has a unique (weak sense) solution. Let $\mathcal{M}_0$ denote the subclass of controls (stationary controls) which do not depend on t (for use in the next section). Let $u(y^\epsilon,\cdot)$, $\bar{u}^\delta(\hat{x}^\epsilon,\hat{z}^\epsilon,\cdot)$ and $\bar{u}^\delta(\hat{x},\hat{z},\cdot)$ denote the controls with values $u(y^\epsilon(\cdot),t)$ $\bar{u}^\delta(\hat{x}^\epsilon(t),\hat{z}^\epsilon(t),t)$ and $\bar{u}^\delta(\hat{x}(t),\hat{z}(t),t)$ at time t.

Theorem 8.1. *Assume the conditions above in this section. For* $\delta > 0$, *let there exist a control* $\bar{u}^\delta(\cdot,\cdot)$ *in* $\mathcal{M}$ *which is* δ*-optimal for* (8.6), (8.7), (8.9), *with respect to controls in* $\bar{\mathcal{H}}$. *Then, for any* $u(\cdot,\cdot) \in \bar{\mathcal{H}}$,

$$(8.10) \qquad \underline{\lim}_{\epsilon} R^\epsilon(u(y^\epsilon,\cdot)) \geq \lim_{\epsilon} R^\epsilon(\bar{u}^\delta(\hat{x}^\epsilon, \hat{z}^\epsilon, \cdot)) - \delta$$
$$= R(\bar{u}^\delta(\hat{x}, \hat{z}, \cdot)) - \delta .$$

Remark. It would be preferable if we could allow the comparison control $u(y^\epsilon(\cdot),\cdot)$ to depend on ϵ other than only via the values of $y^\epsilon(\cdot)$; i.e., for it to be a (say) δ-optimal admissible control for the '*physical*' ϵ*-system*. This is possible, if we can a-priori guarantee some smoothness (uniformly in ϵ) of the obtained controls - so that a weak convergence argument can be carried out - yielding an admissible limit control for the filtering/control problem. But, in general, the limit of $\{u^\epsilon(y^\epsilon(\cdot),\cdot)\}$ would not necessarily be dependent only on the limit data y - even if $y^\epsilon(\cdot) \Rightarrow y(\cdot)$. This is clear from the examples in Section 2. Similar difficulties occur in all work concerning the existence of optimal controls under 'partial information'.

Extensions. The theorem can be carried over to the case where the observations (of both x^ϵ and z^ϵ) are of the non-linear form (6.1), and to the conditional Gaussian case.

Theorem 8.1 can be readily extended to the non-linear case where $\dot{x}^\epsilon = b(x^\epsilon,u) + \sigma(x,\xi_x^\epsilon)$ and $\{x^\epsilon(\cdot)\}$ converges weakly to an appropriate diffusion for 'nice' controls, and where $x^\epsilon(t)$ can be observed without additive noise. If the noise term $\sigma(x,\xi_x^\epsilon)$ were of the control dependent form $\sigma(x,u,\xi_x^\epsilon)$, then there might not be a weak convergence result - unless $u(\cdot)$ were 'smooth'. In the 'smooth' case, there might be a correction term which depended on certain *derivatives* of the control! See Section 5 for additional comment on this point.

9. Filtering and Control: The Large Time Case.

We now treat the filtering and control analog of the large time and bandwidth problem of Section 4, and will use the assumptions

C9.1. $\begin{bmatrix} A_x & 0 \\ 0 & A_z \end{bmatrix} \equiv A$ *is stable,* $[A;H_x,H_y]$ *is observable and* $\begin{bmatrix} A, & \begin{matrix} B_x \\ B_z \end{matrix} \end{bmatrix}$ *controllable.*

C9.2. $\xi^\epsilon(\cdot)$ *satisfies* (C4.2).

The cost functions are

$$\gamma^\epsilon(u) = \overline{\lim_T} \frac{1}{T} \int_0^T E\, r(z^\epsilon(t),x^\epsilon(t),u(t))dt \tag{9.1}$$

$$\gamma(u) = \overline{\lim_T} \frac{1}{T} \int_0^T E\, r(z(t),x^\epsilon(t),u(t))dt \tag{9.2}$$

We adapt the point of view of [18, Section 6] and assume that the system can be Markovianized. This is incorporated in the following assumption.

C9.3. *For each* $\epsilon > 0$, *there is a random process* $\psi^\epsilon(\cdot)$ *such that* $\{\psi^\epsilon(t),\ t < \infty\}$ *is tight and for each* $u(\cdot) \in \mathcal{M}_0$, ($\mathcal{M}_0$ *defined above Theorem* 8.1) $X^\epsilon(\cdot) \equiv \{x^\epsilon(\cdot),z^\epsilon(\cdot),\hat{x}^\epsilon(\cdot),\hat{z}^\epsilon(\cdot),\psi^\epsilon(\cdot),\xi^\epsilon(\cdot),\xi^\epsilon_x(\cdot)\}$ *is a right continuous homogeneous Markov-Feller process (with left hand limits).*

Remark. If $z^\epsilon(\cdot)$ satisfies $\dot{z}^\epsilon = A_z z^\epsilon + \xi_z^\epsilon$, then the assumption (C9.3) holds if the driving noises $(\xi_x^\epsilon(\cdot),\xi_y^\epsilon(\cdot),\xi_z^\epsilon(\cdot))$ satisfy (C9.3) and (C9.1), (C9.2) hold; i.e., if the noises $\xi_z^\epsilon(\cdot)$ and $\xi^\epsilon(\cdot)$ can be written as functions of a suitable Markov process. Let $u(\hat{x}^\epsilon,\hat{z}^\epsilon)$ and $u(\hat{x},\hat{z})$ (and similarly for u^δ) denote controls with values $u(\hat{x}^\epsilon(t),\hat{z}^\epsilon(t))$ and $u(\hat{x}(t),\hat{z}(t))$ at time t.

Theorem 9.1. *Assume the conditions of Theorem* 8.1 *and* (C9.1) - (C9.3). *Let* $\xi^\epsilon(\cdot)$ *and* $\xi_x^\epsilon(\cdot)$ satisfy (C4.2) *and let* $z^\epsilon(\cdot)$ *satisfy* (C4.3). *For* $\delta > 0$, *let there be a* δ*-optimal* control* $\overline{u}^\delta(\cdot,\cdot) \in M_0$ *for the system* (8.1), (8.6), (8.7), *and cost* (9.2), *and for which* (8.1), (8.6), (8.7) *has a unique invariant measure. Then, for* $u(\cdot,\cdot) \in M_o$

$$\text{(9.3)} \qquad \underline{\lim_\epsilon}\, \gamma^\epsilon(u(\hat{x}^\epsilon,\hat{z}^\epsilon)) \geq \lim_\epsilon \gamma^\epsilon(\overline{u}^\delta(\hat{x}^\epsilon,\hat{z}^\epsilon)) - \delta$$

$$= \gamma(\overline{u}^\delta(\hat{x},\hat{z})) - \delta .$$

Proof. Fix $u(\cdot,\cdot) \in M_0$. Define the 'averaged transition measure'

$$P_T^\epsilon(\cdot) = \frac{1}{T} E\int_0^T P\{X^\epsilon(t) \in \cdot \,|X^\epsilon(0)\}dt,$$

where the expectation E is over the possibly random initial conditions, and $X^\epsilon(\cdot)$ is the process corresponding to the use of $u(\hat{x}^\epsilon(\cdot), \hat{z}^\epsilon(\cdot))$. By the

---

*By δ-optimal, one means that it is δ-optimal with respect to all non-anticipative (with respect to the observed data) measurable u-valued controls, for each initial condition.

hypothesis, $\{P^{\epsilon}_{T}(\cdot), T \geq 0\}$ is tight. Also (writing $X = (x,z,\hat{x},\hat{z})$)

$$(9.4) \qquad \gamma^{\epsilon}(u(\hat{x}^{\epsilon},\hat{z}^{\epsilon})) = \overline{\lim_{T}} \int r(x,z,u(\hat{x},\hat{z}))\ P^{\epsilon}_{T}(dX)..$$

Let $\tau^{\epsilon}_{n} \to \infty$ be a sequence such that it attains the limit $\overline{\lim_{T}}$, and for which $P^{\epsilon}_{T^{\epsilon}_{n}}(\cdot)$ converges weakly to a measure, which we denote by $P^{\epsilon}(\cdot)$. Using the 'Feller' property and the right continuity, it is not hard to show that $P^{\epsilon}(\cdot)$ is an invariant measure for $X^{\epsilon}(\cdot)$. Also, by construction of $P^{\epsilon}(\cdot)$,

$$\gamma^{\epsilon}(u(\hat{x}^{\epsilon},\hat{z}^{\epsilon})) = \int \gamma(x,z,u(\hat{x},\hat{z}))P^{\epsilon}(dX) \ .$$

Let $(x^{\epsilon}_{0}(\cdot),z^{\epsilon}_{0}(\cdot),\hat{x}^{\epsilon}_{0}(\cdot),\hat{z}^{\epsilon}_{0}(\cdot))$ denote the first four components of the *stationary Markov-Feller* $X^{\epsilon}(\cdot)$*-process* associated with the invariant measure $P^{\epsilon}(\cdot)$. By our hypotheses (see the argument in Section 4) $\{x^{\epsilon}_{0}(\cdot),z^{\epsilon}_{0}(\cdot),\hat{x}^{\epsilon}_{0}(\cdot),\hat{z}^{\epsilon}_{0}(\cdot)\}$ converges weakly to a limit $(x_{0}(\cdot),z_{0}(\cdot),\hat{x}_{0}(\cdot),\hat{z}_{0}(\cdot))$ satisfying (8.7), (8.1), (8.6). Also, the limit must be stationary, since the $(x^{\epsilon}_{0}(\cdot),...,\hat{z}^{\epsilon}_{0}(\cdot))$ is for each ϵ. Let $\mu^{u}(\cdot)$ denote the invariant measure associated with *this stationary limit*. Then

$$\gamma^{\epsilon}(u(\hat{x}^{\epsilon},\hat{z}^{\epsilon})) \to \gamma(u(\hat{x},\hat{z})) = \int r(x,z,u(\hat{x},\hat{z}))\ \mu^{u}(dxdzd\hat{x}d\hat{z}) \ .$$

By a similar argument, it can be shown that

$$\gamma(\bar{u}^{\delta}(\hat{x},\hat{z})) \equiv \int r(x,z,\bar{u}^{\delta}(\hat{x},\hat{z}))\ \mu^{\bar{u}^{\delta}}\ (dxdzd\hat{x}d\hat{z})$$

$$= \lim_{\epsilon}\ \gamma^{\epsilon}(\bar{u}^{\delta}(\hat{x}^{\epsilon},\hat{z}^{\epsilon})) \ .$$

(The uniqueness of the invariant measure $\mu^{\bar{u}^\delta}(\cdot)$ is used here). Inequality (9.3) now follows from the δ-optimality of $\bar{u}^\delta(\cdot)$. Q.E.D.

Extensions. As for the case of Theorem 8.1, we do not know how to work with arbitrary admissible $u^\epsilon(\cdot)$ as comparison controls. But Theorem 9.1 can be extended in many ways. Perhaps the simplest is the following. For arbitrary q and $t_i \geq 0$, let $u(t)$ depend on $(y(t-t_i) - y(t), i \leq q)$ or (for the ϵ-system) on $(y^\epsilon(t-t_i) - y^\epsilon(t), i \leq q)$ as appropriate, as well as on $\hat{x}(t), \hat{z}(t)$, or on $\hat{x}^\epsilon(t), \hat{z}^\epsilon(t)$, and enlarge the class M_0 to include such dependencies. Then the theorem remains true. More generally, we can allow $u(\cdot)$ to depend on other functionals of the data, provided that those functionals, together with $X^\epsilon(\cdot)$ can be 'appropriately Markovianized' - so that the scheme of the proof can be used, and the uniqueness and non-anticipitative properties continue to hold.

Appendix 1, Weak Convergence Definitions.

Let $P_n(\cdot)$ be the measure associated with X_n, a Euclidean r-space (R^r) valued random variable. We say that $\{X_n\}$ or $\{P_n\}$ is *tight* (equivalently, $\{X_n\}$ is bounded in probability) if $\sup_n P_n\{|X_n| \geq N\} \to 0$ as $N \to \infty$. If $\{X_n\}$ is tight, then by the Helley-Bray Theorem, there is a subsequence $\{n_i\}$ and a measure $P(\cdot)$ and associated random variable X such that $X_{n_i} \to X$ in distribution. Equivalently, $Ef(X_{n_i}) \to Ef(X)$ for each bounded and continuous function. In fact, $f(\cdot)$ can be any bounded measurable function for which $P\{x: f(\cdot)$ discontinuous at $x\} = 0$ ([3], Theorem 5.1). As seen in the text, this is a useful generalization.

Let $C[0,\infty)$ denote the space of continuous functions on $[0,\infty)$ with values in E^r (we always omit the r-dependence in the notation), with the topology of uniform convergence on each bounded interval. The metric on $C[0,\infty)$ can be taken to be

$$d(x(\cdot),y(\cdot)) = \int_0^\infty e^{-t} \max\,[1, \sup_{s\leq t} |x(s) - y(s)|]dt \quad .$$

Let $D[0,\infty)$ denote the space of (R^r-valued) functions on $[0,\infty)$ which are right continuous and have left hand limits, and with the Skorohod topology. See [3], [4] for a discussion of this topology. The topology can be metrized so that the spce is complete and separable. *If* $x(\cdot)$ *is continuous, then* $x_n(\cdot) \to x(\cdot)$ *in this topology if and only if the convergence is uniform on each bounded interval.* This is all that we need to know here. If $d_T(\cdot,\cdot)$ is the metric on $D[0,T]$, then (as above), the metric on $D[0,\infty)$ can be taken to be $\int_0^\infty e^{-t} d_t(x(\cdot),y(\cdot))dt$. The spaces $C[0,\infty)$ and $D[0,\infty)$ are the two most useful (currently) spaces for the study of the convergence of a sequence of random processer. Even if the paths are continuous, it is often more convenient to work with $D[0,\infty)$.

Let $P_n(\cdot)$ be a measure on $D[0,\infty)$ associated with a random process $x_n(\cdot)$ (which we call X_n) whose paths are in $D[0,\infty)$ w.p.1. We say that $P_n(\cdot)$ *converges weakly* (written $\Rightarrow$) to a measure $P(\cdot)$ associated with a process $X = x(\cdot)$ with paths in $D[0,\infty)$ if $Ef(X_n) \to Ef(X)$ for each bounded continuous function $f(\cdot)$ on $D[0,\infty)$. We might also write $X_n \Rightarrow X$. If there is weak convergence, then $f(\cdot)$ can be any measurable function which is continuous only almost everywhere with respect to the limit measure $P(\cdot)$ [3, Theorem 5.1]. The sequence $\{X_n\}$ or $\{P_n\}$ is said to be *tight* if for each $\delta > 0$, there is a compact set $K_\delta \in D[0,\infty)$ such that $P_n\{X_n \in K_\delta\} \geq 1-\delta$ for all n. If $\{X_n\}$ is tight, then there is a subsequence n_i and a $P(\cdot)$ on $D[0,\infty)$ (with associated process $X = x(\cdot)$) such that $X_{n_i} \Rightarrow X$. Analogous definitions and facts hold for processes with paths in $C[0,\infty)$.

There are many useful criteria for tightness and for identifying the limits. For purposes of analysis, it is often useful to alter the probability space so that there is a stronger type of convergence. The choice of the probability space does not affect the weak convergence result - since the disributions of the X_n never changes.

Skorohod imbedding (sometimes called Skorohod representation) [1], [19]. *Let* $P_n \Rightarrow P$ *on* $D[0,\infty)$ (*or on* $C[0,\infty)$). *There is a probability space* $(\tilde{\Omega},\tilde{B},\tilde{P})$ *with processes* $\tilde{X}_n,\tilde{X}$ *defined on it so that* $\tilde{P}\{\tilde{X}_n \in A\} = P\{X_n \in A\}$,

$\tilde{P}\{\tilde{X} \in A\} = P\{X \in A\}$ *for any Borel set* $A \in D[0,\infty)$ (*or in* $C[0,\infty)$, *if we are working in this space) and* $d(\tilde{X}_n,\tilde{X}) \to 0$ w.p.1. Thus, if we wish, we can alter the probability space so that we get w.p.1. convergence in the metric of $D[0,\infty)$ (or $C[0,\infty)$), without altering the distributions of each process X_n or X. This device often facilitates the analysis.

Appendix 2. Proof of Lemma 2.1.

Proof. Choose a finite partition $G = (G_0, G_1, \ldots)$ of R^g such that

$$\text{(A2.1)} \qquad P\{Y \in \partial G_i\} = 0, \text{ all } i; \; P\{Y \in G_i\} > 0, \; i > 0,$$
$$P\{Y \in G_0\} = 0 .$$

(For notational simplicity we omit G_0 below.) Let $\mathcal{F}$ (resp. $\mathcal{F}_n$) denote the σ-algebra on Ω induced by $\{I_{G_i}(Y), i > 0)\}$. (resp., $\{I_{G_i}(Y_n), i > 0\}$). Given $\delta > 0$, we can choose the partition such that

$$\text{(A2.2)} \qquad E[E(X|Y) - E(X|\mathcal{F})]^2 \leq \delta .$$

By Jensen's inequality,

$$\text{(A2.3)} \qquad E[E^2(X_n|\mathcal{F}_n)] \leq E[E^2(X_n|Y_n)] .$$

By (A2.3) and the arbitrariness of δ, to prove the lemma, we need only show that

$$\overline{\lim_n}\; E[X_n - E(X_n|\mathcal{F}_n)]^2$$
$$= \overline{\lim_n}\; [EX_n^2 - EE^2(X_n|\mathcal{F}_n)]$$
$$\leq E[X - E(X|\mathcal{F})]^2 = EX^2 - E[E^2(X|\mathcal{F})],$$

or, equivalently, (since $EX_n^2 \to EX^2$) that

(A2.4) $$\frac{\lim}{n} E\, E^2(X_n | \mathcal{F}_n) \geq E\, E^2(X | \mathcal{F}) \, ,$$

(A2.4) follows from Bayes' rule, the weak convergence and Fatous' lemma since

$$\frac{\lim}{n} E\, E^2(X_n | \mathcal{F}_n) =$$

$$\frac{\lim}{n} \sum_1^\infty \frac{E^2 X_n I_{Y_n}(G_i)}{P(Y_n \in G_i)}$$

$$\geq \sum_1^\infty \frac{E^2 X I_Y(G_i)}{P(Y \in G_i)} = E\, E^2(X | \mathcal{F}) \, .$$

[We used the fact (see Appendix 1) that if $P\{Y \in \partial G_i\} = 0$, then $X_n I_{Y_n}(G_i) \Rightarrow X I_Y(G_i)$.

Q.E.D.

Appendix 3. A Method for Getting Weak Convergence.

In this section, we outline a method for showing that a sequence of solutions to a wide bandwidth noise driven ODE converges weakly to a diffusion, and identify the diffusion. The method is taken from [1, Chapter 5], and is a slight simplification of the method in [8].

Let $x^\epsilon(\cdot)$ be defined by

$$\text{(A3.1)} \qquad \dot{x}^\epsilon = K(x^\epsilon) + F(x^\epsilon)\xi(t/\epsilon^2)/\epsilon,$$

where $\xi(\cdot)$ is a second order stationary right continuous process with left hand limits and integrable correlation function $\bar{R}(\cdot)$, and the functions $K(\cdot)$ and $F(\cdot)$ are continuous, (A3.1) has a unique solution and $F(\cdot)$ is continuously differentiable. Define $\bar{R}_0 = \int_{-\infty}^{\infty} E\xi(u)\xi'(0)du$, assume that

$$\text{(A3.2)} \qquad E\left|\int_s^t du \left[E(\xi(u)\xi'(s) \mid \xi(\tau),\tau\leq 0) - \bar{R}(u-s)\right]\right| \to 0$$

$$\text{as} \quad t,s \to \infty.$$

The condition is not very restrictive. We use it here only because it allows the use of a convenient reference.

Define the diffusion operator $\mathfrak{L}$ and function $\bar{G} = (\bar{G}_1,\ldots)$ by

$$\text{(A3.3)} \qquad \mathfrak{L}f(x) = f_x'(x)\ K(x) + \int_0^\infty E[f_x'(x)F(x)\xi(t)]_x'\ F(x)\xi(0)dt$$

$$\equiv \sum_i f_{x_i}(x)\ \bar{G}_i(x) + \tfrac{1}{2}\ \text{trace}\ \{f_{x_ix_j}(x)\} \cdot \{F(x)\bar{R}_0F'(x)\},$$

where $(\bar{G}_1, \cdots) = \bar{G}$ are the coefficients of the first derivatives $(f_{x_1}, \cdots)$ in (A3.3).

The operator $\mathfrak{L}$ is the differential generator of the Itô process

$$dx = \bar{G}(x)dt + F(x)\bar{R}_0^{\frac{1}{2}}\, dw, \tag{A3.4}$$

where $w(\cdot)$ is a standard Wiener process. Suppose that (A3.5) has a unique solution in the sense of distributions. Then, by [1, Chapter 5.8.4], if $x^\epsilon(0) \Rightarrow x(0)$, then $x^\epsilon(\cdot) \Rightarrow x(\cdot)$, with initial condition $x(0)$.

Appendix 4. A Weak Infinitesimal Operator $\hat{A}^{\epsilon}$.

Refer to the notation of Section 6. Let $f(\cdot)$, $g(\cdot)$ be real valued (progressively) measurable functions of $z^{\epsilon}(\cdot)$, $\xi_y^{\epsilon}(\cdot)$ and let E_t^{ϵ} denote the expectation conditioned on $z^{\epsilon}(s), \xi_y^{\epsilon}(s), s \leq t$. Define the operator $\hat{A}^{\epsilon}$ by: $f(\cdot) \in$ domain of $\hat{A}^{\epsilon}$ and $\hat{A}^{\epsilon} f = g$ if for each T,

$$\lim_{\Delta \to 0} \sup_{\epsilon, t \leq T} E \left| \frac{E_t^{\epsilon} f(t+\Delta) - f(t)}{\Delta} \right| < \infty$$

$$\sup_{t \leq T} E |g(t)| < \infty$$

$$E \left| \frac{E_t^{\epsilon} f(t+\Delta) - f(t)}{\Delta} - g(t) \right| \xrightarrow{\Delta} 0, \quad \text{each } t .$$

Then [1], [8], [16], [17], for $s \geq 0$, $t \geq 0$,

$$E_t^{\epsilon} f(t+s) - f(t) = \int_t^{t+s} E_t^{\epsilon} \hat{A}^{\epsilon} f(u) du. \tag{A4.1}$$

The $\hat{A}^{\epsilon}$ operator plays the role of an infinitesimal operator for non-Markov processes. The relationship (A4.1) has many applications (see the references) in weak convergence theory.

References

[1] H.J. Kushner, *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory*, M.I.T. Press, Cambridge, U.S.A., 1984.

[2] H.J. Kushner and Hai Huang, "Approximate and Limit Results for Nonlinear Filters with Wide Bandwidth Observation Noise", Stochastics, Feb., 1986, also LCDS Rept. 84-36, Brown University.

[3] P. Billingsley, *Convergence of Probability Measures*, 1968, Wiley, New York.

[4] T.G. Kurtz, *Approximation of Population Processes*, 1981, Vol. 36 in CBMS-NSF Regional Conf. Series in Appl. Math, Soc. for Ind. and Appl. Math, Phila.

[5] A. Benveniste, "Design of Monostep and Multistep Adaptive Algorithms for the Tracking of Time Varying Systems," Proc., 23 Conf. on Dec. and Control, 1984, IEEE Publications, New York.

[6] M. El-Ansary and H. Khalil, "On the Interplay of Singular Perturbations and Wide-band Stochastic Fluctuations", SIAM J. on Control, *24*, 1986, 83-98.

[7] H.J. Kushner and Hai Huang, "Averaging Methods for the Asymptotic Analysis of Learning and Adaptive Systems with Small Adjustment Rate", SIAM J. on Control and Optim., *19*, (1981), 635-650.

[8] H. Kushner, "Jump Diffusion Approximations for Ordinary Differential Equations with Wideband Random Right Hand Sides", SIAM J. on Control and Optimization, *17*, 1979, 729-744.

[9] J.M.C. Clark, "The Design of Robust Approximations to the Stochastic Differential Equations of Non-linear Filtering", Ed. J.K. Skwirzynski, NATO Advanced Study Institute Series, Sijthoff and Noordhoff, Alphen aan der Rijn, 1978.

[10] M.H.A. Davis, "A Pathwise Solution of the Equations of Non-linear Filtering", Theory of Prob. and its Applications.

[11] A. Bensoussan and G.L. Blankenship, "Singular Perturbations in Stochastic Control", Univ. of Maryland, E.E. Dept., April, 1985.

[12] R.S. Lipster and A.N. Shiryaev, *Statistics of Random Processes*, 1977, Springer, Berlin.

[13] W.B. Davenport, "Signal to Noise Ratios in Band Pass Limiters", J. Appl. Phys., *24*, 1953, 720-727.

[14] H.J. Kushner, "Diffusion Approximations to Output Processes of Nonlinear Systems with Wide-band Inputs, and Applications", IEEE Trans. on Inf. Theory, *IT-26*, 1980, 715-725.

[15] G.B. Blankenship and G.C. Papanicolaou, "Stability and Control of Stochastic Systems with Wide Band Noise Disturbances", SIAM J. Appl. Math *34*, 1978, 437-476.

[16] T.G. Kurtz, "Semigroups of Conditioned Shifts and Approximations of Markov Processes", Ann. Prob., *4*, 1975, 618-642.

[17] R. Rishel, "Necessary and Sufficient Dynamic Programming Conditions for Continuous Time Stochastic Optimal Control", SIAM J. Control, *8*, 1970, 559-571.

[18] H. Kushner and W. Runggaldier, "Nearly Optimal State Feedback Controls for Stochastic Systems with Wideband Noise Disturbances", to appear SIAM J. on Control and Optimization. Also, LCDS Rept. #85-23, 1985, Brown Univ.

[19] A.V. Skorohod, "Limit Theorems for Stochastic Processes", Theory of Probability and Its Applications, *1*, 1956, 262-290.

END

FILMED

5-86

DTIC